BASE 26

PHILIP M. COHEN
Aliquippa, Pennsylvania

In "Sea-Changed Words" in the February 1969 issue of Word Ways, Howard Bergerson discussed ways of transforming words using numerical equivalents for their letters. In "Ten Logotopian Lingos (Part 2)" in the February 1971 issue of Word Ways, J. A. Lindon presented a conversation in Numbo-Carrean, a language where all words correspond to squares (thus, "doctor" is doc'ddd = 4-15-3-4-4-4 = 4153444 = the square of 2038). However, neither article exploited the possibilities of the most natural link between words and numbers: base 26.

For those unfamiliar with the new math, a brief note is necessary. The ordinary decimal system is "base 10" because we express numbers with powers of ten: thus, $106_{10} = 1x10^2 + 0x10^1 + 6x10^0$ ($10^0 = 1$). The subscript following 106 shows the base. But there is no mathematical reason why we should pick 10; if we chose 26 for a base, 106_{26} would be $1x26^2 + 0x26^1 + 6x26^0 = 682_{10}$, and we would write 106_{10} as $42_{26} = 4x26^1 + 2x26^0$.

We need 26 digits in base 26, not just the 10 digits of base 10, and of course the letters just fill the bill; we let A represent 0, B 1 and so on to Z as 25. For the rest of this article, digits will mean base 10 and capital letters base 26. Thus, BAG = 682 (as we demonstrated above), and PNEUMONOULTRAMICROSCOPICSILICOVOLCANOCONIOSIS is obviously 28142158824471292390070969851145590570537370-73616514570246479290.

Given this correspondence, what can we do with it? Well, let's look for number-words with special properties. ADD, ER, HO, AIR, ALD, WE, BAA, BEE, CAR, CWM, TOR, UPO, CHEM, CRAM, ESSE, FEAR, FINK, HELD, HUNK, LAKE, MOUE, NUCE, VELO, and IRIDO- are square numbers. BB, TS, COM, EBO, AULS, and CRAM are cubes. The names FRYE and BEPPO and the almost-word RISKYY are fifth powers. ABC, CO, DA, EB, HI, AIX, LO, ANN, OO, BIT, CUT, DEN, EON, AERO, AGIO, HEP, HUB, ICA, NET, PAN, WAG, BIRK, BORN, CHAT, COJA, D-CUP, DOLD, DOYT, FEOD, FLOX, GOBO, HEAP, HINA, HOOK, NANA, OBAN, POMO, RIPA, SOHO, TWIG, WEBB and FEKED are triangular numbers (the nth triangular number is the sum of the first n integers -- 1, 3, 6, 10, 15, 21, etc.). Finally, FO, ON, KAF and BAFF are Fibonacci numbers.

Prime words (those which correspond to prime numbers in base 10) are relatively easy to find. The following list is restricted to three-letter words which are found in Webster's Pocket Dictionary:

adz	cap	dub	for	hat	lay	pal	sub
aid	cub	dud	fox	hit	let	pat	sup
apt	cur	eel	fur	hod	lev	pep	vex
art	dab	ell	gab	jar	lox	rev	vip
bar	did	err	gar	job	nil	rib	yip
box	dip	esp	gob	kid	nip	sip	zip
bid	dot	fit	god	lap	oft	sop	

The squares above make up only a small part of the vocabulary of the Base 26 dialect of Numbo-Carrean, of course. While my fluency, vocabulary, and industry are not up to J. A. Lindon's, I have prepared a short Base 26 text dealing with the capture of a thief.

Numbo-Carrean

"Daamb!" Ayb cyrz'd. "A cop-VW 'z heeyr. Eie knno e-muz bee heeyr f'-mee. Aae, dume."

Jeemm Diugn ambyld txo dd abohd. "Eie 'mm Ofzce Diugn. Yewe bee akudzd of-rob ol-Myz Rozaa. Comq outw ahmhye, scom, bayfo eie comq insie afddo yewe. Yewe heeyr?"

"Ah-Dugn, f'-whie? As-sugn az eie 'mm oot-o' heeyr, wam! 'M-syz 'dd. Yewe 'llw hov-to bseej dd abohd f'-daiz!"

"Ho, ih-gez eie hov-to kilw yewe."

"Eaaaa, yewe odyiz fink! Eie 'llw nottk dye-fo' a itie qrkqq. Fyir nottk!"

English

"Damn!" Abe cursed. "A poliçe car is here. I know he must be here for me. Oy, I'm doomed."

Jim Dugan ambled up to the house. "I'm Officer Dugan. You are accused of robbing old Miss Rosa. Come out with your hands up, scum, before I come in after you. You hear?"

"Aw, Dugan, what for? As soon as I'm out of here, pow! I'm caught. I can hole up in this house for days!"

"Hmph, I guess I'll have to kill you."

"Aaagghh, you dirty rat! I won't die for a teeny little peccadillo! Don't shoot!"

Suppose that, following Bergerson, we look for interesting base 26 transformations of words. About the only method with any chance of yielding words from words is multiplication, though it rarely works for words over four letters. Some of the more interesting examples: DIE x 8 = BANG, EAT x 2 = IBM, FOX x 4 = WHO, FUN x 7 = BONN (sounds like a tourist come-on). GNU x 8 goes to CAGE, GUY x 6

arises AB OVO. MAN x 6 changes to CUDA, PIT x 2 rises to a BERM, and when you're OLD x 6, ADIOS. The most productive word I've found is DAD: x 2 = GAG, x 4 = MAM, x 5 = PAP, x 7 = VAV, and x 10 = BEBE. CAB is not bad, either: x 3 = GAD, x 5 = KAF, and x 6 = MAG.

Because of the limited range of transformations purely in base 26, I've concentrated on those that link it to the more familiar base 10.

The first I tried (I played with it by hand for several years before writing a computer program to do the work) may be called reshaping, and works like this: convert the word to its base 10 number, reverse the number (for example, 1737 to 7371), and then cancel these steps by converting the reversed number back to base 26 letters and reversing these.

I had hoped that repeated reshapings would lead almost any word to some other word, but alas, not so -- the blasted things grow. After 99 reshapings, WORD has become RLWGYTOHPSLHJUWTFK. However, a few interesting results do appear. Two reshapings of ONE give FOO, two of THREE give SHI IS, five of NINE give TRY SF (a message from the gods?), five of LOA give TUN, eight of HAP give PAH and ten give POD, and fourteen reshapings of CHESS give EF and a fifteenth, RIB. SEA turns to an appropriate near-word, ZGOB; HE to FAB; NOSE to SLEY; IN to SE, then OK, then FS. ME goes to PX, then via junk (CM, UB) to VE and DC. LEG goes to GEL (via JOJ), BAD to TIC (via OLB).

Analogously, one can decimally behead a word: convert to base 10, cut off the first digit, then back to base 26. Thus, CUP beheads to BID. Decimal curtailment of KEY gives BAM, and OLLA gives BLOC.

The trouble with all these is that in the results all 26 letters are about equally likely to show up, contrary to English letter frequencies. If a method produces only one result -- or even ten results -- for a given word, it will usually produce only garbage. Let us turn, then, to a method that produces 24 or more results per word: permutation.

We hash a word as follows: convert it to base 10, produce all permutations of the digits of the base 10 number, and convert these all back to base 26 letters. If a word gives 4 different decimal digits (the fewest we will consider here), there will be 24 hashes. 5 digits give 120, and 6 digits, 720; thus, as words grow sparser, the number of hashes increases to compensate. If digits repeat, there will be fewer different hashes, of course, and less chance of new words; thus, GOY = 4444 gives only GOY, and RAT = 11511 gives only RAT and four garbages. But generally it works. Of 174 4-digit words tested, only 18 had no hashes in Webster's Second or Third), and all but 2 had hashes that could be found in other sources. Of 32 5-digit words, all but RAT and RED had hashes in Webster's, and RED yielded FEWD, which is in the Oxford English Dictionary. The three 6-digit words tried (six is about the limit for my program) were even richer: WORD hashes to BEGUN, LETE, SAIN, SAWS, WOOS; WAYS hashes to BUNCH and

TOLY; and PLAY to ABUV, AZER, AZIM, BOKIE, GHAT, SAC, SUN and ZER.

Other interesting hashes are DAD/MA (they also hash to IF and ET), ACRE/LOT (also ALSO, ACYL, NEZ), LAY/GAL (also GAU, SI, JOW), LET/INN, HOW/MAD, NET/CUT, GET/NOW, and OLD/FOP. We find that a NEW OAR sweeps clean, as does a NEW MOP, not to mention ANEW, OON and the recent abbreviation NPG. A hashed OWL is a CROC, as well as a plant (ZEA) and BOH, GAPS and GS. A sentence for fishermen is NOT COD, BUB. The richest group is ARP, ASH, GAB, GAT, HAS, ICK, ILK, LAO, LEA, WC.

When scanning through hashes for words, one is constantly frustrated to see one with the right letters for a word, but reversed, like BREV from SIX, or just a bit interchanged, like EMTI and FOSG from FOUR. This leads naturally to the supreme transformation of rehashing: after forming all the hashes, form all possible transposals of each. Here at last we have a method that almost always can derive from a given word some other word (at least, I'd guess, up to 7 or 8 letters). For instance, only two of the 174 4-digit words had no rehashes in Webster's, and even those two can be disposed of using other sources. Among the 5-digit words, only RAT held out. And in the 6-digit case -- well, it's El Dorado. WORD has nearly fifty thousand rehashes, at least 91 of which are words (not counting ones like Butz and Decca which aren't in Webster's). They range from adle to Zulu, starting with all letters but Q and X: Berne, blink, block, boks, cheap, deal, guiac, lieu, peach, plum, sled, Vedic, Whig, ... And WAYS rehashes to at least 57 words.

I wouldn't be surprised if there existed a 5-letter word that could be turned into any other 5-letter (or shorter) word in a couple dozen rehashes. Anyone got a few hours of computer time so I can see? I am willing to lend my PL/I program for reshaping (adaptable to other Base 26 operations) to anyone interested.